WHAT IS CLAIMED IS:

1. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for speech synthesis, the method steps comprising:

determining prosodic parameters of a spoken utterance;

automatically generating a marked-up text corresponding to the spoken utterance using the prosodic parameters; and

generating a synthetic waveform using the marked-up text.

2. The program storage device of claim 1, wherein the instructions for determining prosodic parameters comprise instructions for determining pitch contour, duration contour or energy contour information of the spoken utterance, or any combination thereof.

3. The program storage device of claim 1, further comprising instructions for aligning the spoken utterance with a corresponding text string.

4. The program storage device of claim 3, wherein the instructions for aligning comprise instructions for extracting acoustic feature data from the spoken utterance

and time-aligning the spoken input to the corresponding text string using the acoustic feature data.

5. The program storage device of claim 3, wherein the alignment is performed using Viterbi alignment process.

6. The program storage device of claim 3, wherein the alignment is performed on a phoneme level.

7. The program storage device of claim 1, wherein the instructions for automatically generating a marked-up text comprise instruction for directly specifying the prosodic parameters as attribute values for mark-up elements.

8. The program storage device of claim 1, wherein the instructions for automatically generating a marked-up text comprise instructions for assigning abstract labels to the prosodic parameters to generate a high-level markup.

9. The program storage device of claim 1, wherein the marked-up text is generated using SSML (speech synthesis markup language).

10. The program storage device of claim 1, further comprising instruction for processing phonetic content of the spoken utterance to generate the synthetic waveform having a desired pronunciation.

11. A method for speech synthesis, comprising the steps of:

determining prosodic parameters of a spoken utterance;

automatically generating a marked-up text corresponding to the spoken utterance using the prosodic parameters; and

generating a synthetic waveform using the marked-up text.

12. The method of claim 11, wherein the determining prosodic parameters comprises determining pitch contour, duration contour or energy contour information of the spoken utterance, or any combination thereof.

13. The method of claim 11, further comprising aligning the spoken utterance with a corresponding text string.

14. The method of claim 13, wherein aligning comprises extracting acoustic feature data from the spoken utterance and time-aligning the spoken input to the corresponding text string using the acoustic feature data.

15. The method of claim 13, wherein aligning is performed using Viterbi alignment process.

16. The method of claim 13, wherein aligning is performed on a phoneme level.

17. The method of claim 11, wherein automatically generating a marked-up text comprises directly specifying the prosodic parameters as attribute values for mark-up elements.

18. The method of claim 11, wherein automatically generating a marked-up text comprises assigning abstract labels to the prosodic parameters to generate a high-level markup.

19. The method of claim 11, wherein the marked-up text is generated using SSML (speech synthesis markup language).

20. The method of claim 11, further comprising processing phonetic content of the spoken utterance to generate the synthetic waveform having a desired pronunciation.

21. A text-to-speech (TTS) system, comprising:

a prosody analyzer for determining prosodic parameters of a spoken utterance and automatically generating a marked-up text corresponding to the spoken utterance using the prosodic parameters; and

a TTS system for generating a synthetic waveform using the marked-up text.

22. The system of claim 21, further comprising a user interface that enables a user to input the spoken utterance and input a text string corresponding to the spoken utterance.

23. The system of claim 21, wherein the prosody analyzer processes phonetic content of the spoken utterance to generate the synthetic waveform having a desired pronunciation.

24. The system of claim 21, wherein the prosody analyzer comprises:

a pitch contour extraction module for determining pitch contour information for the spoken utterance;

an alignment module for aligning the input text string with the spoken utterance to determine duration contour information of elements comprising the input text string; and

a conversion module for including markup in the input text string in accordance with the duration and pitch contour information to generate the marked up text.